Estimation methods for multivariate dynamic models

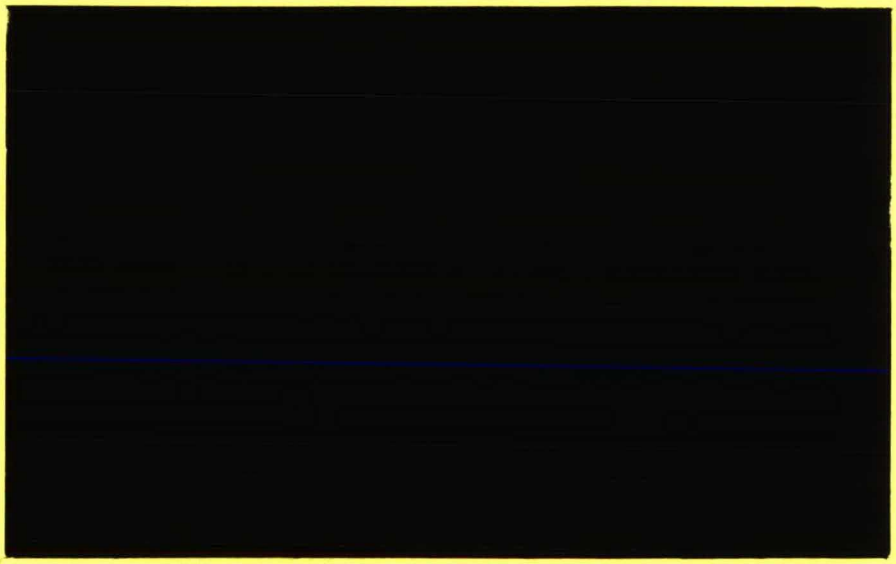
Frijns, J.M.G.

Publication date:
1976

Document Version
Publisher's PDF, also known as Version of record

KATHOLIEKE HOGESCHOOL TILBURG

REEKS "TER DISCUSSIE"



FACULTEIT DER ECONOMISCHE WETENSCHAPPEN

KATHOLIEKE HOGESCHOOL TILBURG

REEKS "TER DISCUSSIE"

No. 76.029 oktober 1976

Estimation methods for multivariate dynamic models

J.M.G. Frijns

FACULTEIT DER ECONOMISCHE WETENSCHAPPEN

1. Modelspecification

In this section we will analyse how a given structural form spcification of a model can be rewritten in other, basically equivalent, specifications. Starting point is the structural form of the model

$$
(1.1)\quad \begin{cases} B(L)Y_t = A(L)X_t + E_t \qquad t = 1,2,\ldots \\ \\ \text{where} \\ B(L) = B_0 + B_1L + \ldots + B_r L^r;\ B_0,\ldots,B_r \text{ are } k\times k \text{ matrices}, \\ \qquad\qquad B_0 \text{ is non-singular.} \\ A(L) = A_0 + A_1L + \ldots + A_p L^s;\ A_0,\ldots,A_s \text{ are } k\times m \text{ matrices}. \\ |A(z)| = 0 \text{ and } |B(z)| = 0 \text{ have no common roots.} \\ \text{All roots of } |B(z)| = 0 \text{ lie outside the unit circle.} \\ E_t \text{ is a widely stationary stochastic process with mean zero.} \end{cases}
$$

The model defined in (1.1) can be interpreted as a simultaneous equations system with lagged dependent variables or if $B_0 = I$ as a multivariate system of difference equations. If $B_0 \neq I$ we obtain the reduced form specification by premultiplying the structural form (1.1) with the matrix B_0^{-1}; if $B_0 = I$ the structural form (1.1) is identical to the reduced form. In our further analysis we will always assume that $B_0 = I$ so that (1.1) defines the reduced form.

The widely stationary stochastic process $\{E_t\}$ can be generated by an AR, MA of ARMA scheme of finite order. We assume that $\{E_t\}$ is defined by an ARMA-scheme

$$
(1.2)\qquad P(L)E_t = Q(L)U_t
$$

where $\{U_t\}$ is a white noise process, $P(L)$ and $Q(L)$ are matrix polynomial equations in the lag operator L of order p and q. The assumption that all roots of $|P(z)| = 0$ lie outside the unit circle guarantees that $\{E_t\}$ is a widely stationary stochastic process. Further we assume in order (to make (statistical) identification possible or) to avoid redundancy in the model specification, that $B(L)$, $A(L)$ and $Q(L)$ have no common

roots and that Q(L) and P(L) have no common roots. We note that a necessary condition for the identification of the autoregressive structure of the Y_t process and the autoregressive structure of the E_t process is the presence of exogenous variables.

In (1.1) we have defined a scheme which generates a stochastic process $\{Y_t\}$. One might ask if there is a stochastic process $\{Y_t\}$ which is a solution of (1.1) and under which conditions such a solution is unique. Assuming fixed initial values $Y_0, Y_{-1}, \ldots, Y_{1-r}$ we obtain from (1.1) by succesive substituting

$$
(1.3)\quad \begin{cases} Y_t = B_t^{+}(L)\, A(L).X_t + B_t^{\&}(L)Y_0 + B_t^{+}(L)E_t \qquad t = 1,2,\ldots \\ \\ \text{where} \\ \\ B_t^{+} = \sum_{\tau=0}^{t-1} (C^{\tau})_{11} L^{\tau} \\ \\ B_t^{\&} = (C^{t})_{11} + (C^{t})_{12} L + \ldots + (C^{t})_{1r}\, L^{r-1} \\ \\ \text{and } (C^{\tau})_{1j} \text{ are determined from the algorithm, described on page 9.} \end{cases}
$$

For given fixed unitial values the stochastic process $\{Y_t\}$ defined in

(1.3) solves (1.1) and is uniquely determined [1].

The stochastic process $\{Y_t, t = 1,2,\ldots\}$ defined in (1.3) is non-stationary; however if the sequence $\{X_t, t = 1,2,\ldots\}$ is uniformly bounded. e.g. $|X_t| < C$ for all t, and if $\{E_t\}$ is a widely stationary stochastic process we can find constants C_1 and C_2 such that

$$|E[Y_t]| < C_1 \qquad \text{for all } t$$

$$\mathrm{Var}[Y_t] < C_1 \qquad \text{for all } t$$

and further we can show that

$$\lim_{t\to\infty} \mathrm{Var}[Y_t]$$

and

$$\lim_{t\to\infty} E[(Y_t - E[Y_t])(Y_{t+s} - E[Y_{t+s}])] \quad \text{exist.}$$

To prove these results we have to use the stationarity assumption of the $\{E_t\}$ process and the property that the variance of Y_t is not affected by the bounded sequence $\{X_t\}$.

We can replace the fixed initial conditions $Y_0, Y_{-1},\ldots,Y_{1-r}$ in (1.3) by the assumption of stochastic initial conditions, e.g. by

---

1) $\{Y_t\}$ is an unique solution of (1.1) if for every other solution $\{Y_t^*, t = 1,2,\ldots\}$ with the same fixed initial values;

$$E[Y_t - Y_t^*]^2 = 0 \quad \text{or if } Y_t = Y_t^* \text{ a.e. for } t = 1,2,\ldots$$

The uniqueness of the solution can now be shown as follows: let $\{Y_t', t = 1,2,\ldots\}$ and $\{Y_t'', t = 1,2,\ldots\}$ be solutions of (1.1) with given fixed initial values and define $Y_t^* = Y_t' - Y_t''$ then follows from substituting $\{Y_t'\}$ and $\{Y_t''\}$ in (1.3) and subtracting both results

$$Y_t^* = 0 \quad \text{a.e. for } t = 1,2,\ldots$$

arbitrary random variables Y_0, $Y_{-1},\ldots,Y_{1-r}$ with finite variances. A slight generalisation is then possible if we assume that $Y_0,Y_{-1},\ldots,Y_{1-r}$ belong to a stochastic process $\{Y_t,\ t = \ldots-1,0,1,\ldots\}$ so that the proces $\{Y_t,\ t = 0,-1,-2,\ldots\}$ is widely stationary.[2)]

The approach in (1.3) is similar to the so-called final form solution of Theil and Boot of a simultaneous equations system with lagged endogenous variables. The form defined in (1.3) will therefore be called the <u>final form</u> specification of the model.

A specification which has gained some popularity in macro-economic models [3)] is the specification of the model in the form of <u>final equations</u>. By suitable transformations we find, if (1.1) is not a block-recursive system [4)].

$$(1.4)\quad \begin{cases} |B(L)|Y_t = b(L)\,A(L)X_t + b(L)E_t & t = s+1,s+2,\ldots \\ s = (m-1)r \\ Y_0,Y_{-1},\ldots,Y_{1-r} \text{ are fixed initial values} \\ Y_1,\ldots,Y_{s-1} \text{ are determined from (1.3)} \\ B^{-1}(L) = b(L)/|B(L)|;\ b(L) \text{ is the adjoint matrix and } |B(L)| \\ \text{is the determinant of } B(L). \end{cases}$$

The unique solution of (1.4) is the stochastic process $\{Y_t|t = s+1,s+2,\ldots\}$ defined in (1.3). Zhe autoregressive coefficients of the endogenous

---

2) See e.g. B.B. v.d. Genugten, WS V, 1976.
v.d. Genugten shows under which conditions with respect to the scheme (1.1) an unique solution $\{Y_t,\ t = \ldots-1,0,1,\ldots\}$ exists.

3) See Tinbergen (1939).

4) See K. Wallis (1975). The term block recursivity now applies to the matrix equation B(L) and not only to B_0 as is usual in the theory of simultaneous equations systems.

variables are identical for all equations in the system (1.4).

If B(L) is block recursive we can write (1.1) as

$$(1.5)\qquad \begin{bmatrix} B_{11}(L) & 0 \\ B_{21}(L) & B_{22}(L) \end{bmatrix} \begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} A_1(L) \\ A_2(L) \end{bmatrix} X_t + \begin{bmatrix} E_{1t} \\ E_{2t} \end{bmatrix}$$

where Y_{1t} is a $k_1 \times 1$ vector and Y_{2t} is a $(k-k_1)\times 1$ vector. The corresponding final equations can be found in Wallis If the matrix polynomial B(L) is block-recursive, specification (1.5) would imply common factors in both sides of (1.5) and thus redundancy.

In addition to the basic structural form specification (1.1) we can define slightly different structural forms; e.g. let the structural form of a model be given by

$$(1.6)\quad B(L)Y_t = D_t^+(L)\, A(L)X_t + E_t \qquad t = 1,2,\ldots$$

where

B(L), A(L) and E_t are defined in (1.1)

$D_t^+(L) = \sum_{\tau=0}^{t-1} \Lambda^\tau L^\tau$ where Λ is a diagonal matrix with elements λ_i, so that $|\lambda_i| < 1$ for $i = 1,\ldots,k$.

The roots of $|A(z)| = 0$ are not equal to λ_i^{-1}, $i = 1,\ldots,k$.

The unique solution of (1.6) is the stochastic process $\{Y_t,\ t = 1,2,\ldots\}$ so that

$$(1.7)\qquad Y_t = B^+(L)D^+(L)A(L)X_t + B_t^{\otimes}(L)Y_0 + B^+(L)E_t \qquad t = 1,2,\ldots.$$

for given fixed initial values $Y_0, Y_{-1}, \ldots, Y_{1-r}$, and $B_t^+(L), B_t^{\otimes}(L)$ are defined in (1.3).

Writing $D(L) = I - \Lambda L$ we can define the <u>super reduced</u> form corresponding to (1.6)

$$(1.8) \qquad D(L)B(L)Y_t = A(L)X_t + D(L)E_t \qquad t = 2,3,4,....$$

where $\begin{cases} Y_0, Y_{-1},....,Y_{1-r} \text{ are fixed initial values} \\ Y_1 \text{ is determined in (1.7)} \end{cases}$

Specification (1.6) and (1.8) are equivalent in the sense that the unique solution of (1.8) is the stochastic process $\{Y_t, t = 2,3,...\}$ defined in (1.7)

The structural form specification (1.6) defines the same "infinite" lag structure for all exogenous variables in one equation and, in principle, different lag structures for each equation. If we wish to define different "infinite" lag structures for all or some variables in one equation we can define the structural form

$$(1.9) \quad \begin{cases} B(L)Y_t = D^+_{1t}(L)A_1(L)X_{1t} + ... + D^+_{mt}(L)A_m(L)X_{mt} + E_t \qquad t = 1,2,... \\ \text{where} \\ B(L) \text{ and } E_t \text{ are defined in (1.1)} \\ A_j(L) \text{ is a k-variate vector polynomial equation, } j = 1,...,m. \\ D^+_{jt} = \sum_{\tau=0}^{t+1} \Lambda_j^\tau L^\tau \text{ where } \Lambda_j \text{ is a diagonal matrix with elements } \lambda_{ji} \\ \text{so that } |\lambda_{ji}| < 1 \text{ for all i,j.} \\ \text{The roots of } |A_j(z)| = 0 \text{ are not equal to } \lambda_{ji}^{-1}, \ i = 1,....,k \text{ en } j = 1,...,m \end{cases}$$

For given initial values $Y_0, Y_{-1},...,Y_{1-r}$ the unique solution of (1.9) is the stochastic process $\{Y_t, t = 1,2,...\}$ given by

$$(1.10) \quad Y_t = B^+_t(L)D^+_{1t}(L)A_1(L)X_{1t}+...+B^+(L)D_{mt}(L)A_m(L)X_{mt}+B^{\otimes}_t(L)Y_0+B^+_t(L)E_t, \ t = 1,2,...$$

where $B^+_t(L)$ and $B^{\otimes}_t(L)$ are defined in (1.3).

Analogously to (1.8) we can define the super reduced form as, writing $D_j(L) = I - \Lambda_j L$,

$$(1.11)\begin{cases} D_m(L)D_{m-1}(L)\ldots D_1(L)B(L)Y_t = D_m(L)\ldots D_2(L)A_1(L)X_{1t} + \ldots \\ + D_{m-1}(L)\ldots D_1(L)A_m(L)X_{mt} + D_m(L)\ldots D_1(L)E_t \qquad t = m+1,m+2,\ldots \\ \text{where } Y_0, Y_{-1}, \ldots, Y_{1-r} \text{ are fixed initial values} \\ \qquad Y_1, \ldots, Y_m \text{ are determined in (1.10)} \end{cases}$$

The specifications in this Section are written in matrix polynomial equations of lag operators, which is natationally convenient but sometimes difficult to interpret. To illustrate the definitions given in this Section we will give some examples where these definitions and corresponding transformations are used.

Example 1

Let the structural form be given by

$$(1.12)\begin{cases} B(L)Y_t = D_t^+(L)A(L)X_t + E_t \qquad t = 1,2,\ldots \\ \text{where} \\ B(L) = I + B\,L;\ B = \begin{bmatrix} -\beta_1 & 0 \\ 0 & -\beta_2 \end{bmatrix},\ |\beta_j| < 1 \quad j = 1,2, \\ A(L) = A \\ D_t^+(L) = \sum_{\tau=0}^{t-1} \Lambda^\tau L^\tau,\ \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix};\ |\lambda_i| < 1 \quad , \quad i = 1,2,\ldots \end{cases}$$

or

$$(1.13)\quad \begin{bmatrix} Y_{1t} - \beta_1 Y_{1t-1} \\ Y_{2t} - \beta_2 Y_{2t-2} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} a_{1i} \sum_{\tau=0}^{t-1} \lambda_1^\tau X_{i,t-\tau} \\ \sum a_{2i} \sum \lambda_2^\tau X_{i,t-\tau} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \qquad t = 1,2$$

By suitable transformations we obtain the super reduced form

$$(1.14)\quad \begin{cases} Y_{1t}-(\lambda_1+\beta_1)Y_{1t-1} + \lambda_1\beta_1\, Y_{1t-2} = \Sigma a_{1i}\, X_{i,t} + \varepsilon_1 t - \lambda_1\varepsilon_{1t-1} \\ Y_{2t}-(\lambda_2+\beta_2)Y_{2t-1} + \lambda_2\beta_2\, Y_{2t-2} = \Sigma a_{2i}\, X_{i,t} + \varepsilon_{2t} - \lambda_1\varepsilon_{2t-1} \\ \text{for } t = 2,3,4,\ldots \end{cases}$$

By succesive subsitutions we obtain the final form from (1.12) for given initial values Y_{10}, Y_{20}

$$(1.15)\quad \begin{bmatrix} Y_{1t} \\ \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} a_{1i} \sum_{\tau=0}^{t-1} \sum_{r=0}^{\tau} \lambda_1^r \beta_1^{\tau-r} X_{i,t-\tau} \\ \\ \Sigma a_{2i} \;\Sigma\; \Sigma\; \lambda_2^r \beta^{\tau-r} X_{i,t-\tau} \end{bmatrix} + \begin{bmatrix} \beta_1^t Y_{10} \\ \\ \beta_2^t Y_{20} \end{bmatrix} + \begin{bmatrix} \sum_{\tau=0}^{t-1} \beta_1^\tau \varepsilon_{1,t-\tau} \\ \\ \sum_{\tau=0}^{t-1} \beta_2^\tau \varepsilon_{2,t-\tau} \end{bmatrix}$$

Example 2

Let the structural form be given by

$$(1.16)\quad \begin{cases} B(L)Y_t = A(L)X_t + E_t \qquad\qquad t = 1,2,\ldots \\ \text{where} \\ B(L) = I + BL,\ B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \text{ B is a neg. def. matrix with char.} \\ \text{roots } \gamma_i, \text{ where } |\gamma_i| < 1,\ i = 1,2. \\ A(L) = A \end{cases}$$

We can write (1.16) as

$$(1.17)\quad \begin{bmatrix} Y_{1t} + b_{11} Y_{1,t-1} + b_{12} Y_{2,t-1} \\ Y_{2t} + b_{21} Y_{2,t-1} + b_{22} Y_{2,t-1} \end{bmatrix} = \begin{bmatrix} \Sigma a_{1i} X_{1,t} \\ \Sigma a_{2i} X_{2,t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad t = 1,2,\ldots$$

The final equations of model (1.16) can be found after some transformations[5] and is given by

$$(1.18)\quad \begin{bmatrix} Y_{1t} + (b_{11} + b_{22})Y_{1,t-1} - (b_{11}b_{22} - b_{12}b_{21})Y_{1,t-2} \\ Y_{2t} + (b_{11} + b_{22})Y_{2,t-1} - (b_{11}b_{22} - b_{12}b_{21})Y_{2,t-2} \end{bmatrix} =$$

$$\begin{bmatrix} \Sigma a_{1i}X_{it} + \Sigma(b_{22}a_{1i} - b_{12}a_{2i})X_{i,t-1} \\ \Sigma a_{2i}X_{it} + \Sigma(b_{11}a_{2a} - b_{21}a_{1i})X_{i,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} + b_{22}\,\varepsilon_{1,t-1} - b_{12}\,\varepsilon_{2,t-1} \\ \varepsilon_{2t} + b_{11}\,\varepsilon_{2,t-1} - b_{21}\,\varepsilon_{1,t-1} \end{bmatrix}$$

for $t = 2,3,\ldots$

Finally we can find the final form by applying the algorithm to be described below.

The algorithm referred to in (1.3) is as follows. Let

$$(1.19)\quad B(L)Y_t = X_t + E_t$$

where $B(L)$ and E_t satisfy the assumptions made in (1.1). Then we can rewrite this r-th order system of difference equations in a, formally equivalent,

---

5) The final equations can be found by subtracting the form

$$\begin{bmatrix} -b_{22} & b_{12} \\ b_{21} & -b_{11} \end{bmatrix} \begin{bmatrix} Y_{t-1} + B\,Y_{t-2} \end{bmatrix}$$

from (1.16) or (1.17).

system of first order difference equations

$$(1.20)\quad \begin{bmatrix} Y_{1t} \\ Y_{2t} \\ \vdots \\ Y_{rt} \end{bmatrix} - \begin{bmatrix} B_1 & \cdots\cdots & B_r \\ I & & 0 \\ & & \\ & I & 0 \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ \vdots \\ Y_{rt-1} \end{bmatrix} = \begin{bmatrix} X_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} E_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where $Y_{1t} = Y_t, \ldots, Y_{rt} = Y_{t-(r-1)}$. System (1.20) can be written in a more compact notation as

$$(1.21)\quad \bar{Y}_t = C\,\bar{Y}_{t-1} + \bar{X}_t + \bar{E}_t$$

where $\bar{Y}_t$, C, $\bar{X}_t$ and $\bar{E}_t$ are defined from (1.20) and (1.21).

By successive substitutions follows from (1.21)

$$(1.22)\quad \bar{Y}_t = \sum_{\tau=0}^{t-1} C^\tau\,\bar{X}_{t-\tau} + C^t\,\bar{Y}_0 + \sum_{\tau=0}^{t-1} C^\tau\,\bar{E}_{t-\tau}$$

and, since we need only the first vector element of $\bar{Y}_t$,

$$(1.23)\quad Y_{1t} = \sum_{\tau=0}^{t-1} (C^\tau)_{11}\,X_{t-\tau} + \sum_{j=1}^{r} (C^t)_{1j}\,Y_{j0} + \sum_{\tau=0}^{t-1} (C^\tau)_{11}\,E_{t-\tau}$$

where $(C^\tau)_{1j}$ is the corresponding submatrix of C^τ and $Y_{1t} = Y_t$, $Y_{10} = Y_0$, $Y_{20} = Y_{-1}, \ldots, Y_{r0} = Y_{1-r}$.

2. Likelihood function and M.L. estimators for the reduced form

2.1. Likelihood function

The model to be analysed can be written as

$$B(L)Y_t = A(L)X_t + \varepsilon_t \qquad t = 1,2,\ldots \tag{2.1}$$

where $\{\varepsilon_t\}$ follows an ARMA-process defined by the scheme

$$P(L)\varepsilon_t = Q(L)u_t \qquad t = 1,2,\ldots \tag{2.2}$$

where $\{u_t\}$ is a multi-variate white noise process with contemporaneous convariance matrix Ω and mean zero and P(L) and Q(L) are matrix polynomial equations of order p and q. Further we assume that the roots of B(L) and P(L) lie outside the unit circle, that B(L) and A(L), have no common roots

and that P(L) and Q(L) have no common roots.[1)]

The super reduced form (S.R.F.) can be seen as a special case of (2.1). Let $B(L) = D(L)B_1(L)$ and $Q(L) = D(L)Q_1(L)$ then (2.1) can be written as

$$(2.3) \qquad D(L)B_1(L)Y_t = A(L)X_t + \varepsilon_t^*$$

where $\qquad P(L)\,\varepsilon_t^* = D(L)Q_1(L)u_t$

so that the difference between the S.R.F. and the R.F. appears from a different specification of the stochastic structure of the error term.

If the R.F. has been obtained from a system of simultaneous equations the parameters of B(L), A(L) and the covariance matrix of Ω are subject to non-linear restrictions which follow from the specification of the structural form and the transformation of structural form to R.F..

---

1) This assumption is stronger than the assumption of no redundancy (Redundancy is the phenomenon that (2.1) is observationally equivalent with the structure $D(L)B(L)Y_t = D(L)A(L)X_t + D(L)\varepsilon_t$ where the roots of $|D(z)| = 0$ are all outside the unit circle. See. eg. M. Hatanaka (1975).) Our assumption guarantees that (2.1) is <u>not</u> obtained by a transformation of the following type: let

(i) $\quad B_1(L) = A_1(L)X_t + P^{-1}(L)Q(L)u_t$

then we can transform (i) to

(ii) $\quad P(L)B_1(L)Y_t = P(L)A_1(L)X_t + Q(L)u_t$

or (iii) $\quad B(L)Y_t = A(L)X_t + \varepsilon_t$

where $B(L) = P(L)B_1(L)$, $A(L) = P(L)A_1(L)$ and $\varepsilon_t = Q(L)u_t$. In our approach (2.1) reflects an underlying economic theory and is not a <u>formal</u> representation of an observable $\{Y_t\}$-process.

If a theoretic foundation of the model specification (i) is lacking it is more convenient to use specification (iii) which shows striking analogies with the well known ARMA-models in the formal prediction models.

See e.g. Koopmans, Rubin and Leipnik (1950, p. 124).

If u_t has a multivariate normal distribution we can write the logarithm of the probability density function of $(u_1,\ldots,u_T)$ as

$$(2.4) \qquad C - \frac{T}{2} \ln|\Omega| - \frac{1}{2} \sum_{t=1}^{T} u_t' \Omega^{-1} u_t$$

From (2.2) and (2.4) follows for given fixed initial values $\varepsilon_0,\varepsilon_{-1},\ldots,\varepsilon_{1-p}$ and $u_0,\ldots,u_{1-q}$ the probability density function of $(\varepsilon_1,\ldots,\varepsilon_T)$. For fixed $(\varepsilon_0,\ldots,\varepsilon_{1-p};\ u_0,\ldots,u_{1-q})$ (2.2) defines an one-to-one transformation of $(u_1,\ldots,u_T)$ into $(\varepsilon_1,\ldots,\varepsilon_T)$. Th Jacobian of this transformation is

$$(2.5) \qquad \left| \left[\frac{\partial u_t}{\partial \varepsilon_s} \right] \right| = \begin{vmatrix} Q_0 & & \\ \vdots & \ddots & \\ \left(\frac{\partial u_t}{\partial \varepsilon_s}\right) & & Q_0 \end{vmatrix} = |Q_0|^T$$

and the logarithm of the conditional probability density function of $(\varepsilon_1,\ldots,\varepsilon_T)$ can be written as

$$(2.6) \qquad C - \frac{T}{2} \ln|\Omega| - T \ln|Q_0| - \frac{1}{2} \sum_{t=1}^{T} (g_t(\varepsilon_t,\varepsilon_{t-1},\ldots,\varepsilon_1,R_{0t}))'$$

$$Q_0^{-1})\ \Omega^{-1}\ Q_0^{-1}(g_t(\varepsilon_t,\varepsilon_{t-1},\ldots,\varepsilon_1,R_{0t}))$$

where $Q_0\ u_t = g_t(\varepsilon_t,\varepsilon_{t-1},\ldots,\varepsilon_1,\ R_{0t})$ and R_{0t} is a function of fixed initial values. The linear vector functions $g_t(.)$ can be (recursively) obtained

from (2.2).[2].

Since (2.1) defines an one-to-one transformation from $(\varepsilon_1,\ldots,\varepsilon_T)$ into $(Y_1,\ldots,Y_T)$ for fixed initial values $(Y_0,\ldots,Y_{1-r})$ we can analogously obtain the probability density function of $(Y_1,\ldots,Y_T)$. The Jacobian of this transformation is 1, since $B_0 = I$, so that the logarithm of the probability density function of $(Y_1,\ldots,Y_T)$ is

$$(2.7) \qquad C - \frac{T}{2}\ln|\Omega| - T\ln|Q_0| - \frac{1}{2}\Sigma(g_t(B(L)Y_t - A(L)X_t,\ldots,R_{0t}))Q_0^{-1})\Omega^{-1}Q_0^{-1}(g_t(.))$$

Given the probability density function of $(Y_1,\ldots,Y_T)$ for given fixed initial values $(Y_0,\ldots,Y_{1-r},\varepsilon_0,\ldots,\varepsilon_{1-p}, u_0,\ldots,u_{1-q})$ and the corresponding log-likelihood function we can obtain ML-estimators for the parameters of B(L) and A(L) and the parameters of P(L), Q(L) and Ω. The number of parameters can be fairly large so that M.L.-estimators for this model are only meaningful for large samples.

If we can approximate the ARMA-process of $\{\varepsilon_t\}$ by a finite order A.R.process or a finite order M.A.process a substantial reduction in the

---

2) It is of course possible to obtain directly the covariance matrix of $(\varepsilon_1,\ldots,\varepsilon_T)$ which are realisations of the ARMA process defined by

$$P(L)\varepsilon_t = Q(L)u_t \qquad t = \ldots,-1,0,1,2,\ldots$$

Let $(\varepsilon_{11},\ldots,\varepsilon_{1k},\ldots,\varepsilon_{T1},\ldots,\varepsilon_{Tk})$ have covariance matrix Σ then the loglikelihood function is proportional with

$$-\frac{1}{2}\ln|\Sigma| - \frac{1}{2}\varepsilon'\,\Sigma^{-1}\varepsilon$$

where $\varepsilon = (\varepsilon_{11},\ldots,\varepsilon_{1k},\ldots,\varepsilon_{T1},\ldots,\varepsilon_{Tk})$. Using this specification we can easily derive the loglikelihood function of $(Y_1,\ldots,Y_t)$ for fixed initial values $Y_0,\ldots,Y_{1-r}$. Since Σ will have in general a very complicated structure, ML estimators based on this structure will require laborious computations, so that for multivariate models at least this approach does not seem desirable. However Kang (undated) suggest that for small samples M.L. estimators based on this specification (which does not depend on given fixed initial values $\varepsilon_0,\ldots,\varepsilon_{1-p}$) are better than ML estimators based on specification (2.7).

number of parameters is possible. Let us firstly analyse the case where the ARMA-process can be approximated by a A.R. process of order p [3]

$$(2.8) \qquad P(L)\varepsilon_t = u_t \qquad t = 1,2,\ldots$$

where $\{u_t\}$ is a multivariate normally distributed white noise process with mean zero and non-singular covariance matrix Ω and $(\varepsilon_0,\ldots,\varepsilon_{1-p})$ are fixed initial values.

Combining (2.8) in (2.1) we obtain

$$(2.9) \qquad P(L)B(L)Y_t = P(L)A(L)X_t + u_t \qquad t = 1,2,\ldots$$

---

3) Let $\{\varepsilon_t\}$ follow from the ARMA scheme

$$P_1(L)\varepsilon_t = Q(L)u_t$$

and assume that the finite order M.A.process $\eta_t = Q(L)u_t$ can be approximated by the infinite M.A.process $\eta_t = P_2^{-1}(L)u_t$ so that the roots of $|P_2(z)| = 0$ lie outside the unit circle. We obtain

$$P_1(L)\varepsilon_t = P_2^{-1}(L)u_t$$

or

$$P_2(L)P_1(L)\varepsilon_t = u_t$$

or $P(L)\varepsilon_t = u_t$ where $P(L) = P_2(L)P_1(L)$.

Second order A.R. processes can be used to describe a wide variety of weight distributions, so that in many practical situations the finite M.A. process can be adequately approximated by a second order A.R. process.

with fixed initial values $(Y_0,\ldots,Y_{1-r-p})$[4]. Since $(u_1,\ldots,u_T)$ are independent multivariate normal variables we can easily obtain the prob. dens. function of $(Y_1,\ldots,Y_T)$. The log likelihood function of $(Y_1,\ldots,Y_T)$ is

$$(2.10)\quad C - \frac{T}{2}\ln|\Omega| - \frac{1}{2}\sum_t (P(L)(B(L)Y_t - A(L)X_t))' \Omega^{-1}(P(L)(B(L)Y_t - A(L)X_t))$$

since the Jacobian of the transformation is 1 if $B_0 = I$ and $P_0 = I$.

Model (2.9) is a special case of the more general model

$$(2.11)\quad B(L)Y_t = A(L)X_t + u_t$$

where $\{u_t\}$ is a white noise process which demonstrates that in many practical cases it will be very difficult to distinguish between the autoregressive structure of the $\{Y_t\}$ process and the autoregressive structure caused by the AR process of the error term ε_t. Identification is only possible if the regression equation contains exogenous variables. See e.g. L. Kenward (1975) or D. Hendry (1975) where the specification (2.9) is tested against the more general case (2.11).

If the stochastic process of $\{\varepsilon_t\}$ is generated by a finite order M.A. scheme

$$(2.12)\quad \varepsilon_t = Q(L)u_t \qquad t = 1,2,\ldots$$

we can write model (2.1) as

$$(2.13)\quad Y_t = (B(L) - I)Y_t + A(L)X_t + Q(L)u_t$$

---

4) The fixed initial values $(Y_{-r},\ldots,Y_{1-r-p})$ follow from

$$\varepsilon_t = B(L)Y_t - A(L)X_t \qquad t = 0,\ldots,1-p$$

where $(\varepsilon_0,\ldots,\varepsilon_{1-p})$ and $(Y_0,\ldots,Y_{1-r})$ are given fixed initial values.

The sample $(Y_1,\ldots,Y_T)$ can then be written in a very compact form as

(2.14) $\quad Y = Y\,\beta + X\,\alpha + \eta + Q.u$

where (2.13) is the t-th row of (2.14) and η is a vector of fixed initial "effects" such that

(2.15)
$$\begin{bmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_2\end{bmatrix} = \eta + \begin{bmatrix} Q_0, & 0 & \\ Q_1, Q_0, & 0, & \\ \vdots & & \\ 0, \ldots\ldots\ldots & 0, Q_q, \ldots\ldots & Q_0\end{bmatrix}\begin{bmatrix}u_1\\ \vdots\\ u_T\end{bmatrix} = \eta + Q\,u$$

We can rewrite (2.14) as

(2.16) $\quad Y = Z\,\delta + Q\,u$

or

(2.17) $\quad u = Q^{-1}\,(Y - Z\,\delta)$

Since u is a vector of independent multivariate normal variables we can easily obtain the log-likelihoodfunction of $Y = (Y_1,\ldots,Y_T)$. The logarithm of the prob. density function of u can be written as

$$C - \frac{T}{2}\ln|\Omega| - \frac{1}{2}\,u'\,(I \otimes \Omega^{-1})u$$

In (2.17) an on-to-one transformation from u into Y is defined with Jacobian $|Q_0|^T$ so that the log-likelihood function of Y is

(2.18) $\quad C - \frac{T}{2}\ln|\Omega| - T\,\ln|Q_0| - \frac{1}{2}\,(Y-Z\delta)'Q^{-1\prime}(I \otimes \Omega^{-1})Q^{-1}(Y-Z\delta)$

Usually Q_0 will be the identity matrix so that $\mathrm{Ln}|Q_0| = 0$, and $Q_1,\ldots,Q_q$ will be diagonal matrices which greatly simplifies the $(k\times T)\times(k\times T)$ matrix Q.

2.2. Maximum likelihood estimators

The computation of M.L. estimators for the model (2.9) is rather laborious. The log likelihood function is a function of the parameters β, α, π, Ω^{-1} corresponding to B(L), A(L), P(L) and the covariancematrix of u_t. We can maximize $L(\beta, \alpha, \pi, \Omega^{-1})$ with respect to Ω^{-1} and then maximize the "concentrated" function $L^c(\beta, \alpha, \pi \,|\hat{\Omega}^{-1})$ with respect to α, β, π. This procedure yields the global maximum of $L(\beta, \alpha, \pi, \Omega^{-1})$[1]. Differentiating (2.10) with respect to Ω^{-1} gives as first order conditions [2]

$$(2.19) \quad \frac{\partial L}{\partial \Omega^{-1}} = \frac{T}{2}\Omega - \frac{1}{2}\sum_t (P(L)(B(L)Y_t - A(L)X_t))(P(L)(B(L)Y_t - A(L)X_t))' = 0$$

which implies that at the maximum

$$(2.20) \quad \hat{\Omega}^{-1} = [\frac{1}{T}\sum_t (P(L)(\ldots))(P(L)(\ldots))']^{-1}$$

and that the concentrated likelihood function can be written as

$$L(\beta, \alpha, \pi|\hat{\Omega}^{-1}) = C - \frac{T}{2} \ln \left|\frac{1}{T}\sum_t (P(L)(\ldots))(P(L)(\ldots))'\right|$$

which is a complicated non-linear function of (β, α, π).

If however P(L) = I and all equations have the same regressors the log likelihood function can be written in the notation of the multivariate regression model as

---

1) In this and subsequent sections it is implicitly assumed that $\Theta = (\alpha, \beta, \pi, \Omega)$ (or the class of prob. distributions P_Θ) is identifiable.

2) See for the differentiation of matrix expressions T.W. Anderson (1958) or H. Theil (1971, p. 31, 32). Important results are $\frac{\partial(y'Bz)}{\partial B} = Y z'$; $\frac{\partial \log|A|}{\partial A} = (A')^{-1}$.

$$L = -\frac{T}{2} \ln |\Omega| - \frac{1}{2} (Y^* - (I_k \otimes Z^*)\delta)'(\Omega^{-1} \otimes I_T)(Y^* - (I_k \otimes Z^*)\delta)$$

where $Y^* = (Y_{11},\ldots,Y_{1T},\ldots,Y_{k1},\ldots,Y_{kT})'$,
$Z^* = (Y_{-1},\ldots,Y_{1-r}, X_{1,1},\ldots,X_{1,1-s},\ldots,X_{m,1},\ldots,X_{m,1-s})$, $X_{i,1} = (X_{i,1},\ldots,X_{i,T})'$ and $\delta = (\beta, \alpha)$. The M.L. estimators of $(\beta, \alpha, \Omega^{-1})$ are then (see T.W. Anderson, (1958, Section 8.2).

$$\hat{\delta} = (Z^{*'} Z^*)^{-1} Z^{*'} Y^*$$

$$\hat{\Omega}^{-1} = [\frac{1}{T} \Sigma(\widehat{B(L)}Y_t - \widehat{A(L)}X_t)(\widehat{B(L)}Y_t - \widehat{A(L)}X_t)']^{-1}$$

A direct iterative procedure to obtain the ML estimates in the general case uses an initial estimate $\hat{\Omega}^{(0)}$ and then maximizes in the first step

$$(2.21) \quad L(\beta, \alpha, \pi|\hat{\Omega}^{(0)}) = C - \frac{T}{2} \ln|\Omega^{(0)}| - \frac{1}{2} \Sigma_t (P(L)(\ldots))'(\hat{\Omega}^{(0)})^{-1}(P(L)(\ldots))$$

with respect to β, α and π, which is equivalent with minimizing the generalized sum of squares in the third term of (2.21). The resulting estimates can be used to compute an estimate $\hat{\Omega}^{(1)}$ using (2.20). In the second step we repeat this procedure replacing $\hat{\Omega}^{(0)}$ in (2.21) by $\hat{\Omega}^{(1)}$ and using the resulting estimates of β, α, π to compute $\hat{\Omega}^{(2)}$. This procedure is continued till convergence occurs. To guarantee that the absolute maximum of the likelihood function is reached we have to repeat this procedure for several initial estimates $\hat{\Omega}^{(0)}$. See e.g. Chow, G.C. (1968).[3)]

A further simplification of the computation procedure is obtained if we use an iterative procedure where Ω and the parameters π of P(L) are in the first step replaced by initial estimates $\hat{\Omega}^{(0)}$ and $\hat{\pi}^{(0)}$. In the first step we have to maximize

$$(2.22) \quad L(\beta, \alpha |\hat{\Omega}^{(0)}, \hat{\pi}^{(0)})$$

---

3) Another possibility is to use the constrained direct search technique (e.g. Box's Complex method (1965)).

which yields $(\hat{\beta}^{(1)}, \hat{\alpha}^{(1)})$ which can be used to compute $\hat{\Omega}^{(1)}$ and $\hat{\pi}^{(1)}$, using the residuals $e_t^{(1)} = \widehat{B(L)}Y_t - \widehat{A(L)}X_t$. The iteration is continued till convergence occurs, and is repeated for several initial estimates $\hat{\Omega}^{(0)}$ and $\hat{\pi}^{(0)}$.

Under certain additional regularity conditions the ML estimators defined in (2.19), (2.21) and (2.22) have desirable asymptotic properties. We can interpret model (2.9) as a special case of the model which is analysed in the study of Koopmans, Rubin and Leipnik (KRL) (1950) on FIML-methods for simultaneous equation models[4]. M.L. estimators of these models are under appropriate regularity conditions consistent, asymptotically normally distributed and asymptotically efficient in the Rao sense (see e.g. P. Schönfeld (1971, Vol II, p. 289).

We can also ignore the prior information on (2.9) so that we obtain the linear model (2.11). In T.W. Anderson(1971, Section 5.4.5.5) it is shown, for the univariate case, that the ML estimators of this model are consistent and asymptotically normally distributed.

The covariance matrix of the asymptotic distribution can be obtained from the likelihood function, see K.R.L, (1950, Section 3.3.10), or P. Schönfeld (1971, Section 18.3.5). A numerical procedure to compute the estimated covariance matrix can be found in S. Schim van der Loeft and R. Harkema (1974).

Analogous to the M.L. estimators for the model (2.9) which is based on an autoregressive model with an error-term which follows a finite order A.R. process we can define M.L. estimators for an autoregressive model with an errorterm which follows a finite order M.A. process. We can also define an iterative procedure to compute the M.L. estimates. This model implies more burdensome computations since it requires the computation

---

4) The model analysed by K.R.L. can be written as

$$Y_t = \Pi\, V_t + U_t$$

subject to (non-linear) restrictions on the parameter matrix Π

$$\psi(\Pi, R) = 0$$

where V_t is vector of lagged endogenous and exogenous variables and $\{u_t\}$ is a multivariate normally distributed white noise process.

of the matrix Q^{-1} in (2.18) in every iteration.

Further it is more difficult to obtain the asymptotic properties of the M.L. estimators. The M.L. estimators of δ in model (2.16) are equivalent with the M.L. estimators of the transformed model

$$(2.23) \quad Y_t = f_t(Y_{t-1},\ldots,Y_1,X_t,\ldots,X_1, \eta, \alpha, \beta, Q^{-1}) + u_t$$

where $f_t(.)$ is a non-linear function of a fixed number of unknown parameters α, β, Q^{-1}, the fixed initial effects η and all (previous) observations $Y_{t-1},\ldots,Y_1, X_t,\ldots,X_1$.

This model differs from the model analysed in the study of K.L,R. (1950) or in Anderson(1971). In a paper presented at the North American Regional Econometric Conference(1966), Phillips A.W.(1966) has shown for the univariate case, and under the assumption that $\{u_t\}$ is a normally distributed white noise process and that certain regularity conditions are satisfied, that application of the M.L. approach will yield consistent and asymptotically efficient estimates of all unknown parameters[5].

---

5) The model (2.23) contains the unobservable begin effect η: for T large we can safely ignore this term. Since the influence of η on the value of the likelihood function converges to zero if $T \to \infty$ it is impossible to obtain as consistent estimators of this effect. See Dhrymes (1971).

2.3. Conclusion

The M.L. estimators defined in this Section require rather complicated computing techniques. In the next Section we will define more-step G.L.S. estimators which are computationally simpler and have the additional advantage that they do not require normality of the errorterm u_t.

A sometimes decisive advantage of the M.L. estimators is their close connection with the likelihood ratio test. Likelihood ratio tests can be used to test different modelspecifications such as (2.9) versus (2.11).

3. More-step G.L.S. estimators for the reduced form

Starting point are the model specifications (2.9) and (2.11). To estimate (2.11) we use a more step procedure. In the first step we apply O.L.S. to (2.11) and use the O.L.S. residuals to compute an estimate of the matrix Ω, $\hat{\Omega}^{(0)}$. In the second step we compute G.L.S. estimates of (α,β) by minimizing

$$(3.1) \qquad \sum_t (B(L)Y_t - D(L)X_t)'(\hat{\Omega}^{(0)})^{-1}(B(L)Y_t - A(L)X_t)$$

with respect to the coefficients α and β of $A(L)$ and $B(L)$[1]. These estimators are, under the modelassumptions with respect to (2.11) and some other usual regularity conditions, consistent and asymptotically normally distributed. In fact the GLS estimators defined in (3.1) are consistent and as. normally distributed for every matrix $\hat{\Omega}^{(0)}$ which converges in probability to a non-singular matrix[2].

The covariance matrix of the asymptotic distribution of the GLS estimators is difficult to obtain and depends on the way an estimate $\hat{\Omega}$ is obtained. See Amemiya and Fuller(1967) or Dhrymes(1971) for some simple cases. If however all equations of model (2.11) contain the same set of regressors, the GLS estimators defined in (3.1) are equivalent with the OLS estimators of each equation seperately. Using lemma 9.2.1 of Schönfeld (1971) we can then easily obtain the covariance matrix of the asymptotic distribution.

---

1) We can also compute non-linear GLS estimators which minimize

$$(3.2) \qquad S(\alpha,\beta,\Omega) = \sum_t (B(L)Y_t - D(L)X_t)'\Omega^{-1}(B(L)Y_t - A(L)X_t)$$

The estimates $\hat{\alpha}$, $\hat{\beta}$, $\hat{\Omega}$ which minimize (3.2) can be found by an iteration process as described in Section (2.2).

2) Writing $\hat{\Omega}^{(0)} = P'P$ the form (3.1) can be written as

$$(3.3) \qquad \sum_t (P\,B(L)Y_t - P\,D(L)X_t)'(P\,B(L)Y_t - PA(L)X_t)$$

Estimators of model (2.9) can be obtained in a similar way as estimators of model (2.11). In the first step we apply unrestricted OLS to model (2.9) and use the OLS residuals to compute an estimate of the matrix Ω, $\hat{\Omega}^{(0)}$. In the second step we compute restricted G.L.S. estimates of (α,β,π) by minimizing

$$(3.4) \qquad \sum_t (P(L)B(L)Y_t - P(L)A(L)X_t)'(\hat{\Omega}^{(0)})^{-1}(P(L)B(L)Y_t - P(L)A(L)X_t)$$

with respect to (α,β,π). Comparing the results of model (2.9) and (2.11) we can test the validity of the restrictions used in (3.4). These tests are analogous to the likelihood ratio tests referred to in Section 2.

In Schönfeld (1971, p. 67) an alternative least squares estimator is suggested for an autoregressive model with autocorrelated errors. Let the model be

$$(3.5) \qquad B(L)Y_t = A(L)X_t + \varepsilon_t$$

$$P(L)\varepsilon_t = u_t$$

In the first step we apply OLS to (3.5) and use the OLS residuals to compute extimates $\hat{\pi}^{(0)}$ of the parameters π of $P(L)$ and an estimated covariance matrix $\hat{\Omega}^{(0)}$ of the multivariate random variable u_t. These estimates are used in the second step where we obtain GLS estimates which minimize

---

continuation note 2

Minimizing (3.3) as function of (α,β) implies that the estimates $(\hat{\alpha},\hat{\beta})$ are the OLS estimates of the transformed model

$$PB(L)Y_t = PA(L)X_t + P\,u_t$$

The OLS estimator for this model, with temporary uncorrelated errors, is constistent and asymptotically normally distributed.

$$(3.6)\qquad \Sigma(\hat{P}^{(0)}(L)B(L)Y_t-\hat{P}^{(0)}(L)A(L)X_t)'(\hat{\Omega}^{(0)})^{-1}$$

$$(\hat{P}^{(0)}(L)B(L)Y_t-\hat{P}^{(0)}(L)A(L)X_t)$$

The residuals $Y_t - \hat{Y}_t$ in this step can be used to obtain estimates $\hat{\pi}^{(1)}$, $\hat{\Omega}^{(1)}$ which can be used in a next iteration. The procedure is continued till convergence occours and can be repeated for several initial values $\hat{\pi}^{(0)}$, $\hat{\Omega}^{(0)}$ to assure that the global minimum is reached. This procedure is analogous to the simplified iterative M.L. procedure described in Section 2.2; its asymptotic properties in the case that u_t has no normal distribution are not clear.

4. Instrumental variables estimators for the reduced form

The use of M.L. or more step G.L.S. estimators in models with lagged dependent variables is only justified if we know the maximum order of the AR process of the error term in advance. Misspecification of the order of this A.R.process implies inconsistent estimates if the order is too low and implies inefficient estimates if the order is too high (See Amermiya and Fuller (1967).

We can avoid the risk of inconsistent estimates due to an underspecification of the order of the error A.R.-process by using I.V.E. (instrumental variable estimators). Let us write the multivariate model as

$$(4.1) \qquad Y = X\delta + \varepsilon$$

where $Y = (Y_{11},\ldots,Y_{1T},\ldots,Y_{k1},\ldots,Y_{kT})'$, $X = [I_k \otimes V]$ where V is a matrix of lagged endogenous variables and exogenous variables and $\delta = (\delta_1,\ldots,\delta_k)'$. The I.V.E. is now defined as (see D. Hendry (1975a) and Sargan (1964)) the vector $\hat{\delta}$ which minimizes

$$(4.2) \qquad (Y - X\delta)' M(Y - X\delta)$$

where $M = Q(Q'Q)^{-1} Q'$; $Q = [I_k \otimes W]$ and W is a matrix of instrumental variables. W consists of the exogenous variables in V plus lagged exogenous variables corresponding to the lagged endogenous variables in V[1].

The first order conditions of (4.2) yield

$$(4.3) \qquad \tilde{\delta} = (X'M\,X)^{-1}(X'M\,Y)$$

if W has been chosen so that (X'M X) is a non-singular matrix.

---

1) If Q = I we obtain the OLS estimator as a special case of the I.V.E.

The estimator $\tilde{\delta}$ can be written as

$$(4.4) \qquad \tilde{\delta} = [I_k \otimes ((V'W)(W'W)^{-1}(W'V))^{-1}((V'W)(W'W)^{-1}W'] \; Y$$

so that

$$(4.5) \qquad \tilde{\delta}_i = (V'W(W'W)^{-1}W'V)^{-1} \; V'W(W'W)^{-1} \; W'Y_i$$

The estimator $\tilde{\delta}$ defined in (4.2) is equivalent with the I.V. estimator which follows from the equations

$$(4.6) \qquad Z'Y = Z'X \; \delta + Z'\varepsilon$$

where $Z = X'Q(Q'Q)^{-1} \; Q'$ so that $Z'X$ is a nonsingular matrix.

The I.V.E. is under general conditions consistent and asymptotically normally distributed [2] though asymptotically less efficient than the G.L.S. estimator based on the true correlation structure of the error term. See eg. Sargan(1964) and Dhrymes (1971).

We can also define a generalized instrumental vairables estimator (G.I.V.E.) which is based on information about the contemporaneous covariance matrix of ε. Let Cov $(\varepsilon_t \; \varepsilon_t')$ = Ω then we define the G.I.V.E. as the vector $\tilde{\tilde{\delta}}$ which minimizes [3].

---

2) Let Plim $\frac{1}{T}$ X'Q be a non-stochastic matrix then under very general conditions we obtain

$$\text{Plim } \frac{1}{T} Z'\varepsilon = \text{plim } \frac{1}{T} X'Q \text{ plim } T(Q'Q)^{-1} \text{ plim } \frac{1}{T} Q'\varepsilon = 0$$

which implies the consistency of $\tilde{\delta}$.

To prove asymptotic normalily we need additional assumptions with respect to the moments of ε.

3) If Q = I we obtain

$$(Y-X\delta)(\Omega^{-1} \otimes I_T)(Y-X\delta)$$

which yields the G.L.S. estimators for the multivariate model with contemporaneously correlated errors with covariance matrix Ω.

$$(4.7)\quad \begin{cases} (Y - X\delta)'N(Y-X\delta) \\ \\ \text{where} \\ \\ N = (\Omega^{-1} \otimes I_T)Q[Q'(\Omega^{-1} \otimes I_T)Q]^{-1} Q'(\Omega^{-1} \otimes I_T) \end{cases}$$

The G.I.V.E. can also be interpreted as the I.V.E. of the transformed model, using $\Omega^{-1} = R'R$,

$$(4.8)\quad Y^* = X^* \delta + \varepsilon^*$$

with corresponding matrix of instrumental variables Q^* and $M^* = Q^*(Q^{*'} Q^*)^{-1} Q^{*'}$ where $Y^* = PY$, $X^* = PY$, $Q^* = PQ$ and

$$(4.9)\quad P'P = [R' \otimes I_T][R \otimes I_T] = [\Omega^{-1} \otimes I_T]$$

From the first order conditions of (4.7) follows

$$(4.10)\quad \tilde{\tilde{\delta}} = (X' NX)^{-1} X'N Y$$

Since $X = [I_k \otimes V]$, $Q = [I_k \otimes W]$ we find after some manipulations

$$(4.11)\quad \tilde{\tilde{\delta}} = [\Omega^{-1} \otimes V'W(W'W)^{-1} W'V]^{-1} [\Omega^{-1} \otimes V'W(W'W)^{-1} W']Y$$

$$= [I_k \otimes (V'W (W'W)^{-1} W'V)^{-1} V'W(W'W)^{-1} W']Y$$

or

$$(4.12)\quad \tilde{\tilde{\delta}}_i = (V'W)(W'W)^{-1} W'V)^{-1} V'W(W'W)^{-1} W'Y_i$$

so that G.I.V.E. and I.V.E. are equivalent for this special case [4].

A disadvantage of I.V. methods is that they are asymptotically less efficient than M.L. or G.L.S. estimators (for correctly specified models). In general I.V. estimators are very useful to provide initial estimates for more-steps estimation procedures.

---

4) If $X \neq \left[I_k \otimes V \right]$ we can still define a matrix Q and a G.I.V.E. analogous to (4.7). These estimators can be shown to be consistent and asymptotically normally distributed, even if we replace the matrix Ω by an estimated matrix $\hat{\Omega}$ so that Plim $\hat{\Omega} = \Omega$, or any other non-stochastic non-singular symmetric matrix.

5. Estimation of the final form

The final form specification is given in (1.3). We can derive the (conditional) likelihood function of $(Y_1,\ldots,Y_T)$ given the fixed initial values. For properly chosen initial values this likelihood function will be equivalent to the likelihood function already obtained in Section 2, equation (2.7). This implies that the M.L. estimates based on the likelihood function of the final form are identical to the M.L. estimates based on the likelihood function of the reduced form.

It is however possible to obtain slightly more general models by assuming

$$(5.1)\quad \begin{cases} Y_t = B^{-1}(L)A(L)X_t + \varepsilon_t & t = 1,2,\ldots \\ \text{with} \\ P(L)\varepsilon_t = Q(L)u_t \end{cases}$$

where $\{u_t\}$ is a multivariate white noise process [1]. If the exogenous variables in (5.1) have no common lag distributions we can define, if there are m regressors,

$$(5.2)\quad \begin{cases} Y_t = \sum_{i=1}^{m} B_i^{-1}(L)A_i(L)X_t + \varepsilon_t \\ \text{with} \\ P(L)\varepsilon_t = Q(L)u_t \end{cases}$$

---

1) We assume again that all roots of $|B(z)| = 0$ and $|P(z)| = 0$ lie outside the unit circle, and that B(L) and A(L), and P(L) and Q(L), have no common roots. See footnote 1, Section 2.1.

From the most general model (5.2) we can arrive at the most restrictive model (1.3) by imposing restrictions on the coefficients of $B_i(L)$, $A_i(L)$, and P(L). Comparing the estimation results of the different models it is possible to test the validity of these restrictions.

D.A.Pierce (1971,1972) derives direct least squares etimators for (the parameters of) the specifications (5.1) and (5.2) for the univariate case. If u_t has a normal distribution these least squares estimators are equivalent with the maximum likelihood estimators. The estimators are under appropriate regularity conditions consistent and asymptotically normally distributed. Further Pierce obtains the covariance matrix of the asymptotic distribution.

The direct least squares estimators require, particularly for multivariate models, laborious non-linear computing methods. Further these direct least squares estimators are highly sensitive for an under-specification of the order of the correlation structure of the error term. In example 1 we will show that for a simple model where the error ε_t follows a second order A.R. process the direct L.S. estimators are inconsistent if we falsely assume that the order of the A.R. process of the error ε_t is one. In this example it is also shown that O.L.S. or more-step G.L.S. will yield consistent estimators of the most important parameters even if the order of the A.R. is misspecified (too low).

Example 1. Let

$$(E.1) \qquad Y_t = \alpha X_t + \varepsilon_t$$

and

$$(E.2) \qquad \varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + u_t$$

where $\{u_t\}$ is a white noise process and the roots of $(1-\rho_1 z-\rho_2 z^2) = 0$ lie outside the unit circle.

Let us wrongly assume that

(E.3) $\varepsilon_t = \rho\ \varepsilon_{t-1} + u_t$

so that we define the direct least squares estimates $(\hat{\alpha}, \hat{\rho})$ as the vector which minimizes

(E.4) $\Sigma(Y_t - \rho\ Y_{t-1} - \alpha\ X_t + \rho\ \alpha\ X_{t-1})^2.$

Thus the direct least squares estimator of (E.1) under assumption (E.3) is equivalent with the non-linear least squares etimator of the non-linear model

(E.5) $Y_t = \rho\ Y_{t-1} + \alpha\ X_t - \rho\ \alpha\ X_{t-1} + \eta_t \qquad \eta_t = \varepsilon_t - \rho\ \varepsilon_{t-1}$

where η_t is wrongly assumed to be serially uncorrelated. If (E.2) holds then η_t is serially correlated so that, since (E.5) contains lagged endogenous variables, the least squares estimator of (E.5) under assumption (E.3) will yield inconsistent estimates of α.

Now let us define a more-step estimation procedure where in the first step OLS is applied to (E.1). Then we find that the OLS estimator $\tilde{\alpha}$ is consistent and moreover that the "estimator" $\tilde{\rho} = (\Sigma\ e_t\ e_{t-1})/\ \Sigma\ e_{t.1}^2$, where e_t are the OLS residuals, has a fixed probability limit. If (E.2) holds we find

(E.6) $\text{plim}\ \hat{\rho} = \sum_{i=0}^{\infty} \rho_2^i\ \rho_1 = \rho_1/(1-\rho_2)$

Using $\hat{\rho}$ we transform, in the usual way, (E.1) to

(E.7) $Y^* = \alpha\ X_t^* + \varepsilon_t^*$

and apply in the second step OLS to the transformed model (E.7) so that

(E.8) $\hat{\alpha}^* = (X^{*\prime}\ X^*)^{-1}\ X^{*\prime}\ Y^* = \alpha + (X^{*\prime}\ X^*)^{-1}(X^{*\prime}\ \varepsilon^*)$

Since $\text{Plim}(\frac{X^{*\prime}\ X^*}{T})^{-1}$ exists and is non-stochastic we conclude that $\text{Plim}\ \hat{\alpha}^* = \alpha$ if $\text{Plim}(\frac{X^{*\prime}\ \varepsilon^*}{T}) = 0$. We obtain

$$(E.9) \quad \text{Plim}\left(\frac{X^{*\prime}\,\varepsilon^{*}}{T}\right) = \text{Plim}\,\frac{1}{T}\,\Sigma\,(X_t - \hat{\rho}\,X_{t-1})(\varepsilon_t - \hat{\rho}\,\varepsilon_{t-1})$$

$$= \text{Plim}\,\frac{X'\varepsilon}{T} - \text{Plim}\,\hat{\rho}\;\;\text{Plim}\,\frac{X'_{-1}\varepsilon}{T}$$

$$- \text{Plim}\,\hat{\rho}\;\text{Plim}\,\frac{X'\,\varepsilon_{-1}}{T} + \text{Plim}\,\hat{\rho}^2\;\text{Plim}\,\frac{X'_{-1}\,\varepsilon_{-1}}{T}$$

$$= 0 + 0 + 0 + 0 = 0.$$

Thus we conclude that the more step G.L.S. estimator is consistent.

We wil concentrate in this Section on two-step G.L.S. estimators which seem more robust then direct (generalized) least squares estimators and which are in general more easy to compute. Let the model be specified in (5.1) We will confine our selves to models where the error ε_t follows a k-th. order A.R. process. To estimate (5.1) we have to use non-linear estimation techniques.

The general form of a multivariate non-linear model is

$$(5.3)\begin{cases} Y_t = g_t(\theta) + \varepsilon_t & \theta \in \Theta \\ \text{with} \\ P(L)\varepsilon_t = u_t \end{cases}$$

where $g_t(\theta)$ is a vector function. We can write the i-th component of this system of equations as

$$(5.4) \quad Y_{it} = g_{it}(\theta_i) + \varepsilon_{it} \qquad \theta_i \in \Theta_i \;\; ; \; i = 1,\ldots,k$$

where $\Theta = \Theta_1 \times \Theta_2 \times \ldots \times \Theta_m$. From (5.1) follow certain restrictions on Θ_i which can be written as

$$(5.5)\quad \begin{cases} R(\theta_1,\ldots,\theta_k) = 0 \\ \text{or as} \\ \theta_1 \in \theta_1^R,\ldots,\theta_k \in \theta_k^r \text{ so that } \theta \in \theta^R \\ \text{where } \theta^R = \theta_1^R \times \ldots \times \theta_k^R. \end{cases}$$

The unrestricted O.L.S. of (5.3) is defined as the vector $\theta \in \theta$ which minimizes

$$(5.6)\quad \sum_t (Y_t - g_t(\theta))'(Y_t - g_t(\theta)) = \sum_{i=1}^{k} \sum_t (Y_{it} - g_{it}(\theta_i))^2$$

The unrestricted O.L.S. of (5.3) is thus identical to unconstrained O.L.S. for the seperate equations. Under certain regularity conditions these estimators are consistent. See Appendix B. The O.L.S. residuals con be used to construct in a second step Feasible Generalized least Squares which are under appropriate regularity conditions consistent and asymptotically normally distributed. See Appendix B for a detaited treatment of F.G.L.S. estimators.

To test the restrictions (5.5) we can use the asymptotic distribution of the F.G.L.S. estimators. A computationally more simple test procedure is based on a comparison of generalized sums of squared residuals under the different hypotheses.[2)]. Though computationally more simple the (asymptotic) properties of this test procedure are more difficult to obtain. In Appendix A, Section 7, we have shown that a test statistic based on the M.L.-residuals is asymptotically equivalent (or at least approximately equal) to a likelihood ratio test which has under appropriate regularity

---

2) Let S_1 be the generalized sum of squares under H_0 and let S_2 be the generalized sum of squares under H_1 ($S_2 \leq S_1$) then the test-statistic is defined as

$$T = (S_1 - S_2)/(S_2/n), \quad n = \text{number of observations}$$

and H_0 is rejected if $T \geq T^*$. To compute T^* we have to determine the (asymptotic) distribution of T_0.

conditions asymptotically a χ^2 distribution. It seems reasonable to expect that the F.G.L.S. estimator defined in App. B is, in the case of a normally distribted u_t, asymptotically equivalent with the M.L.-estimator, in the sense that

$$\text{Plim } \sqrt{n}(\hat{\theta}_{FGLS} - \hat{\theta}_{ML}) = 0.$$

We may thus expect that the asymptotic distribtuion of the test-statistic based on the F.G.L.S.-residuals can be approximated by a χ^2 distribution with the usual number of d.f.

Remark

If the weights of the lagdistribution $B^{-1}(L)A(L)$ in (5.1) are concentrated on small lags, say the first p periods, we can approximate the infinite lag distribution defined by $B^{-1}(L)\ A(L)$ by a finite lag distribution. This finite lag distributions has weights $W_0, W_1, \ldots, W_p$ corresponding to lags of $0,1,\ldots,p$ periods, so that

$$W_i \geq 0, \quad \Sigma W_i = 1 \quad , i = 0,\ldots,p.$$

The assymption of a finite lag distribution reduces the model (5.1) to a linear model

$$(5.7) \qquad Y_t = B_1 X_t + \ldots + B_p X_{t-p} + \varepsilon_t$$

which can be analysed in the context of the general (multivariate) linear model. See eg. P. Dhrymes(1971,Ch 8) or J. Kmenta (1971, Ch 11)

6. Estimation of final equations

Finally we can write the model as a system of final equations, defined in (1.14). The (conditional) likelihood function of $(Y_1,\ldots,Y_T)$ is for properly chosen initial values equivalent to the likelihood function obtained in (2.17). It is however possible to interpret. (1.14) as a special case of the more general model

$$(6.1)\begin{cases} B^D(L)Y_t = A(L)X_t + \varepsilon_t \\ \text{with} \\ P(L)\varepsilon_t = Q(L)u_t \end{cases}$$

where $B^D(L)$ is a diagonal matrix, whose diagonal elements are polynomial equations $B_i(L)$. It is of course assumed that the roots of $|B^D(z)| = 0$ and $|P(z)| = 0$ lie outside the unit circle and that $B_i(L)$ and the corresponding $A_i(L)$ have no common roots (and that $P(L)$ and $Q(L)$ have no common roots).

Estimation of (6.1) requires complicated methods since we have to take in account the correlation structure of ε_t (lagged endogenous variables!). Neglecting the presence of contemporary correlation between ε_{it} and u_{it} $(i = 1,\ldots,k)$ we can estimate each equation of (6.1) seperately. The problem of estimating one equation of (6.1) reduces to the problem of estimating univariate stochastic difference equations. Under certain regularity conditions consistent and asymptotically normally distributed estimators can be obtained. Since we neglect the contemporaneous correlation structure of ε_t these estimators will not be efficient. (These estimates can be used to compute (initial) estimates of the contemporary covariances between ε_{it} (or u_{it}, $i = 1,\ldots,k$) which can be used in a system estimator of (6.1).)

It is thus possible to reduce a system of multivariate difference equations to a system of Seemingly Unrelated Regression equations such that each (lagged) endogenous variable appears in only one regression equation . The advantages of this transformation are not obvious as far as the estimation procedure is concerned.

Appendices

Appendix A and Appendix B will be published as a separate paper in the series "Ter Discussie" of the K.H.T.

Literature

- Amemiya T. and W.A. Fuller, (1967), A comparative study of alternative estimators in a distributed lag model, Econometrica, vol 35.

- Amemiya T., (1973) Generalized least squares with an estimated auto-covariance matrix, Econometrica, Vol 41

- Anderson T.W., (1958), An introduction to multivariate statistical analysis, Wiley, New York.

- Anderson T.W., (1971), Time Series analysis, Wiley, New York

- Box M.J., (1965), A new method of constrained optimization and a comparison with other methods, Computer Journal, Vol 8.

- Chow G.C., (1968), Two methods of computing full-information maximum likelihood estimates in simultaneous stochastic equations, International Economic Review, Vol. 9

- Dhrymes P, (1971), Distributed lags, Holden Day

- v.d. Genugten B.B., (1976), Syllabus WS V, mimeographed, K.H.T.

- Hendry D.F., (1974), Stochastic specification in an aggregate demand model of the United Kingdom, Econometrica, Vol 42

- Hendry D.F. (1975,a), The structure of simultaneous equations estimators, Journal of Econometrics, Vol. 4

- Hendry D.F., (1975,b), Testing dynamic specification in a model of building society behaviour, mimeographed, London School of Economics

- Hendry D.F. and Trivedi P.K. (1972), Maximum likelihood estimation of difference equations with moving average erros: a simulation study, Review of Economic Studies, Vol 39

- Kenward L.R., (1975), Autocorrelation and dynamic methodology with an application to wage determination models, Journal of Econometrics, Vol 3

- Kmenta J., (1971), Elements of Econometrics, Mac Millan, New York

- Koopmans T.C., H. Rubin and R.B. Leipnik (1950), Measuring the equation systems of dynamic economics, in Statistical inference in dynamic economic models, ed. Koopmans T.C., Wiley, New York

- Nicholls D.F, Pagan A.R. and R.D. Terrell (1975), The estimation and use of models with moving average disturbance terms: a survey, International Economic Review, Vol 16

- Newbold P., (1974), The exact likelihood function for a mixed autoregressive-moving average process, Biometrika, Vol 61

- Phillips A.W., (1966), The estimation of systems of difference equations with moving average disturbances (Paper read to the Econometric Society Meeting, San Francisco, 1966)

- Pierce D.A., (1971,a) Distribution of residual autocorrelations in the regression model with autoregressive-moving average errors, Journal of the Royal Statistical Association, Vol 33

- Pierce D.A., (1971,b), Least squares estimation in the regression model with autoregressive-moving average errors, Biometrika, Vol 58

- Pierce D.A., (1972), Least squares estimation in dynamic disturbance time series models, Biometrika, Vol 59

- Sargan J.D., (1964), Wages and prices in the U.K.: a study in Econometric Methodology, in Econometric analysis for National Economic Planning, ed. P.E. Hart et.al., Butterworth scientific publishers

- Schim v.d. Loef S., and R. Harkema, (1974), Three models of firm behaviour; theory and estimation, with an application to the Dutch manufacturing sector, mimeographed, E.U.R.

- Trivedi P.H., (1970) Inventory behaviour in U.K. manufacturing 1956-67, The Review of Economic Studies, Vol 37

- Schönfeld P, (1971), Methoden der Ökonometrie, Band II, Franz Vahlen, Berlin

- Theil H., (1971), Principles of Econometrics, North Holland Publishing Company, Amsterdam

- Tinbergen J, (1939), Statistical testing of business-cycle theories, II; Business cycles in the U.S.A. 1919-1932, Geneva: League of Nations.

- Wallis K.F, (1975), Testing dynamic specification from the final form, paper presented to the Econometric Society World Congress, Toronto, 1975

- Zellner A and F. Palm, (1974), Time series analysis and simultaneous equation econometric models, Journal of Econometrics, Vol 2

In de Reeks ter Discussie zijn verschenen:

1.H.H. Tiggelaar	Spectraalanalyse en stochastische lineaire differentievergelijkingen.	juni '75
2.J.P.C.Kleijnen	De rol van simulatie in de algemene econometrie.	juni '75
3.J.J. Kriens	A stratification procedure for typical auditing problems.	juni '75
4.L.R.J. Westermann	On bounds for Eigenvalues	juni '75
5.W. van Hulst J.Th. van Lieshout	Investment/financial planning with endogenous lifetimes: a heuristic approach to mixed integer programming.	juli '75
6.M.H.C.Paardekooper	Distribution of errors among input and output variables.	augustus '75
7.J.P.C. Kleijnen	Design and analysis of simulation Practical statistical techniques.	augustus '75
8.J. Kriens	Accountantscontrole met behulp van steekproeven.	september '75
9.L.R.J. Westermann	A note on the regula falsi	september '75
10.B.C.J. van Velthoven	Analoge simulatie van ekonomische modellen.	november '75
11.J.P.C. Kleijnen	Het ekonomisch nut van nauwkeurige informatie: simulatie van ondernemingsbeslissingen en informatie.	november '75
12.F.J. Vandamme	Theory change, incompatibility and non-deductibility.	december '75
13.A. van Schaik	De arbeidswaardeleer onderbouwd?	januari '76
14.J.vanLieshout J.Ritzen J.Roemen	Input-ouputanalyse en gelaagde planning.	februari '76
15.J.P.C.Kleijnen	Robustness of multiple ranking procedures: a Monte Carlo experiment illustrating design and analysis techniques.	februari '76
16.J.P.C. Kleijnen	Computers and operations research: a survey.	februari '76
17.J.P.C. Kleijnen	Statistical problems in the simulation of computer systems.	april '76
18.F.J. Vandamme	Towards a more natural deontic logic.	mei '76
19.J.P.C. Kleijnen	Design and analysis of simulation: practical, statistical techniques.	juni '76
20.H.H. Tigelaar	Identifiability in models with lagged variables.	juli '76
21.J.P.C. Kleijnen	Quantile estimation in regenerative simulation: a case study.	augustus '76
22.W.Derks	Inleiding tot econometrische modellen van landen van de E.E.G.	augustus '76
23.B. Diederen Th. Reijs W. Derks	Econometrisch model van België.	september '76
24.J.P.C. Kleijnen	Principles of Economics for computers.	augustus '76
25.B. van Velthoven	Hybriede simulatie van ekonomische modellen.	augustus '76.

26.F. Cole	Forecasting by exponential smoothing, the Box and Jenkins procedure and spectral analysis. A simulation study.	september '76
27.R. Heuts	Some reformulations and extensions in the univariate Box-Jenkins time series analysis.	juli '76
28.W. Derks	Vier econometrische modellen.	